A04

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 108 790 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
20.06.2001 Bulletin 2001/25

(21) Application number: **00127688.0**

(22) Date of filing: **18.12.2000**

(51) Int Cl.[7]: **C12Q 1/68**, C07H 21/04, C12N 15/63, C07K 14/34, C12R 1/15, G06F 17/00, C12R 1/13, G01N 33/50

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **16.12.1999 JP 37748499**
07.04.2000 JP 2000159162
03.08.2000 JP 2000280988

(83) **Declaration under Rule 28(4) EPC (expert solution)**

(71) Applicant: **KYOWA HAKKO KOGYO CO., LTD.**
Chiyoda-ku, Tokyo 100-8185 (JP)

(72) Inventors:
- **Nakagawa, Satochi,**
  c/o Kyowa Hakko Kogyo Co.,Ltd.
  Machida-shi, Tokyo 194-8533 (JP)
- **Mizoguchi, Hiroshi,**
  c/o Kyowa Hakko Kogyo Co.,Ltd.
  Machida-shi, Tokyo 194-8533 (JP)
- **Ando, Seiko, c/o Kyowa Hakko Kogyo Co., Ltd.**
  Machida-shi, Tokyo 194-8533 (JP)
- **Hayashi, Mikiro,**
  c/o Kyowa Hakko Kogyo Co.,Ltd.
  Machida-shi, Tokyo 194-8533 (JP)
- **Ochiai, Keiko, c/o Kyowa Hakko Kogyo Co.,Ltd.**
  Machida-shi, Tokyo 194-8533 (JP)
- **Yokoi, Haruhiko,**
  c/o Kyowa Hakko Kogyo Co.,Ltd.
  Machida-shi, Tokyo 194-8533 (JP)
- **Tateishi, Naoko,**
  c/o Kyowa Hakko Kogyo Co.,Ltd.
  Machida-shi, Tokyo 194-8533 (JP)
- **Senoh, Akihiro, c/o Kyowa Hakko Kogyo Co.,Ltd.**
  Machida-shi, Tokyo 194-8533 (JP)
- **Ikeda, Masato, c/o Kyowa Hakko Kogyo Co.,Ltd.**
  Machida-shi, Tokyo 194-8533 (JP)
- **Ozaki, Akio, c/o Kyowa Hakko Kogyo Co., Ltd.**
  Hofu-shi, Yamaguchi 747-8522 (JP)

(74) Representative: **VOSSIUS & PARTNER**
Siebertstrasse 4
81675 München (DE)

(54) Novel polynucleotides

(57) Novel polynucleotides derived from microorganisms belonging to coryneform bacteria and fragments thereof, polypeptides encoded by the polynucleotides and fragments thereof, polynucleotide arrays comprising the polynucleotides and fragments thereof, recording media in which the nucleotide sequences of the polynucleotide and fragments thereof have been recorded which are readable in a computer, and use of them.

EP 1 108 790 A2

Printed by Jouve, 75001 PARIS (FR)

Description

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0001] The present invention relates to novel polynucleotides derived from microorganisms belonging to coryneform bacteria and fragments thereof, polypeptides encoded by the polynucleotides and fragments thereof, polynucleotide arrays comprising the polynucleotides and fragments thereof, computer readable recording media in which the nucleotide sequences of the polynucleotide and fragments thereof have been recorded, and use of them as well as a method of using the polynucleotide and/or polypeptide sequence information to make comparisons.

2. Brief Description of the Background Art

[0002] Coryneform bacteria are used in producing various useful substances, such as amino acids, nucleic acids, vitamins, saccharides (for example, ribulose), organic acids (for example, pyruvic acid), and analogues of the above-described substances (for example, N-acetylamino acids) and are very useful microorganisms industrially. Many mutants thereof are known.

[0003] For example, *Corynebacterium glutamicum* is a Gram-positive bacterium identified as a glutamic acid-producing bacterium, and many amino acids are produced by mutants thereof. For example, 1,000,000 ton/year of L-glutamic acid which is useful as a seasoning for umami (delicious taste), 250,000 ton/year of L-lysine which is a valuable additive for livestock feeds and the like, and several hundred ton/year or more of other amino acids, such as L-arginine, L-proline, L-glutamine, L-tryptophan, and the like, have been produced in the world (*Nikkei Bio Yearbook 99*, published by Nikkei BP (1998)).

[0004] The production of amino acids by *Corynebacterium glutamicum* is mainly carried out by its mutants (metabolic mutants) which have a mutated metabolic pathway and regulatory systems. In general, an organism is provided with various metabolic regulatory systems so as not to produce more amino acids than it needs. In the biosynthesis of L-lysine, for example, a microorganism belonging to the genus *Corynebacterium* is under such regulation as preventing the excessive production by concerted inhibition by lysine and threonine against the activity of a biosynthesis enzyme common to lysine, threonine and methionine, i.e., an aspartokinase, (*J. Biochem., 65*: 849-859 (1969)). The biosynthesis of arginine is controlled by repressing the expression of its biosynthesis gene by arginine so as not to biosynthesize an excessive amount of arginine (*Microbiology, 142*: 99-108 (1996)). It is considered that these metabolic regulatory mechanisms are deregulated in amino acid-producing mutants. Similarly, the metabolic regulation is deregulated in mutants producing nucleic acids, vitamins, saccharides, organic acids and analogues of the above-described substances so as to improve the productivity of the objective product.

[0005] However, accumulation of basic genetic, biochemical and molecular biological data on coryneform bacteria is insufficient in comparison with *Escherichia coli, Bacillus subtilis,* and the like. Also, few findings have been obtained on mutated genes in amino acid-producing mutants. Thus, there are various mechanisms, which are still unknown, of regulating the growth and metabolism of these microorganisms.

[0006] A chromosomal physical map of *Corynebacterium glutamicum* ATCC 13032 is reported and it is known that its genome size is about 3,100 kb (*Mol. Gen. Genet., 252*: 255-265 (1996)). Calculating on the basis of the usual gene density of bacteria, it is presumed that about 3,000 genes are present in this genome of about 3,100 kb. However, only about 100 genes mainly concerning amino acid biosynthesis genes are known in *Corynebacterium glutamicum,* and the nucleotide sequences of most genes have not been clarified hitherto.

[0007] In recent years, the full nucleotide sequence of the genomes of several microorganisms, such as *Escherichia coli, Mycobacterium tuberculosis,* yeast, and the like, have been determined (*Science, 277*: 1453-62 (1997); *Nature, 393*: 537-544 (1998); *Nature, 387*: 5-105 (1997)). Based on the thus determined full nucleotide sequences, assumption of gene regions and prediction of their function by comparison with the nucleotide sequences of known genes have been carried out. Thus, the functions of a great number of genes have been presumed, without genetic, biochemical or molecular biological experiments.

[0008] In recent years, moreover, techniques for monitoring expression levels of a great number of genes simultaneously or detecting mutations, using DNA chips, DNA arrays or the like in which a partial nucleic acid fragment of a gene or a partial nucleic acid fragment in genomic DNA other than a gene is fixed to a solid support, have been developed. The techniques contribute to the analysis of microorganisms, such as yeasts, *Mycobacterium tuberculosis, Mycobacterium bovis* used in BCG vaccines, and the like (*Science, 278*: 680-686 (1997); *Proc. Natl. Acad. Sci. USA, 96*: 12833-38 (1999); *Science, 284*: 1520-23 (1999)).

(ii) a data storing device for at least temporarily storing the input information;
(iii) a comparator that compares the at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001 with the target amino acid sequence information for determining a function of a polypeptide having the target amino acid sequence which is coincident with or analogous to the polypeptide having at least one amino acid sequence selected from SEQ ID NOS:3502 to 7001; and
(iv) an output device that shows a function obtained by the comparator.

30. A method based on a computer for determining a function of a polypeptide having a target amino acid sequence derived from a coryneform bacterium, comprising the following:

(i) inputting at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001, function information based on the amino acid sequence, and target amino acid sequence information;
(ii) at least temporarily storing said information;
(iii) comparing the at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001 with the target amino acid sequence information; and
(iv) determining a function of a polypeptide having the target amino acid sequence which is coincident with or analogous to the polypeptide having at least one amino acid sequence selected from SEQ ID NOS:3502 to 7001.

31. The system according to any one of claims 23, 25, 27 and 29, wherein a coryneform bacterium is a microorganism of the genus *Corynebacterium*, the genus *Brevibacterium*, or the genus *Microbacterium*.

32. The method according to any one of claims 24, 26, 28 and 30, wherein a coryneform bacterium is a microorganism of the genus *Corynebacterium*, the genus *Brevibacterium*, or the genus *Microbacterium*.

33. The system according to claim 31, wherein the microorganism belonging to the genus *Corynebacterium* is selected from the group consisting of *Corynebacterium glutamicum, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium callunae, Corynebacterium herculis, Corynebacterium lilium, Corynebacterium melassecola, Corynebacterium thermoaminogenes*, and *Corynebacterium ammoniagenes*.

34. The method according to claim 32, wherein the microorganism belonging to the genus *Corynebacterium* is selected from the group consisting of *Corynebacterium glutamicum, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium callunae, Corynebacterium herculis, Corynebacterium lilium, Corynebacterium melassecola, Corynebacterium thermoaminogenes*, and *Corynebacterium ammoniagenes*.

35. A recording medium or storage device which is readable by a computer in which at least one nucleotide sequence information selected from SEQ ID NOS:1 to 3501 or function information based on the nucleotide sequence is recorded, and is usable in the system of claim 23 or 27 or the method of claim 24 or 28.

36. A recording medium or storage device which is readable by a computer in which at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001 or function information based on the amino acid sequence is recorded, and is usable in the system of claim 25 or 29 or the method of claim 26 or 30.

37. The recording medium or storage device according to claim 35 or 36, which is a computer readable recording medium selected from the group consisting of a floppy disc, a hard disc, a magnetic tape, a random access memory (RAM), a read only memory (ROM), a magneto-optic disc (MO), CD-ROM, CD-R, CD-RW, DVD-ROM, DVD-RAM and DVD-RW.

38. A polypeptide having a homoserine dehydrogenase activity, comprising an amino acid sequence in which the Val residue at the 59th in the amino acid sequence of homoserine dehydrogenase derived from a coryneform bacterium is replaced with an amino acid residue other than a Val residue.

39. A polypeptide comprising an amino acid sequence in which the Val residue at the 59th position in the amino acid sequence as represented by SEQ ID NO:6952 is replaced with an amino acid residue other than a Val residue.

40. The polypeptide according to claim 38 or 39, wherein the Val residue at the 59th position is replaced with an Ala residue.

(iv) screening and analyzing nucleotide sequence information which is coincident with or analogous to the target sequence or target structure motif information.

25. A system based on a computer for identifying a target sequence or a target structure motif derived from a coryneform bacterium, comprising the following:

(i) a user input device that inputs at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001, and target sequence or target structure motif information;
(ii) a data storage device for at least temporarily storing the input information;
(iii) a comparator that compares the at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001 with the target sequence or target structure motif information, recorded by the data storage device for screening and analyzing amino acid sequence information which is coincident with or analogous to the target sequence or target structure motif information; and
(iv) an output device that shows a screening or analyzing result obtained by the comparator.

26. A method based on a computer for identifying a target sequence or a target structure motif derived from a coryneform bacterium, comprising the following:

(i) inputting at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001, and target sequence information or target structure motif information into a user input device;
(ii) at least temporarily storing said information;
(iii) comparing the at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001 with the target sequence or target structure motif information; and
(iv) screening and analyzing amino acid sequence information which is coincident with or analogous to the target sequence or target structure motif information.

27. A system based on a computer for determining a function of a polypeptide encoded by a polynucleotide having a target nucleotide sequence derived from a coryneform bacterium, comprising the following:

(i) a user input device that inputs at least one nucleotide sequence information selected from SEQ ID NOS:2 to 3501, function information of a polypeptide encoded by the nucleotide sequence, and target nucleotide sequence information;
(ii) a data storage device for at least temporarily storing the input information;
(iii) a comparator that compares the at least one nucleotide sequence information selected from SEQ ID NOS: 2 to 3501 with the target nucleotide sequence information for determining a function of a polypeptide encoded by a polynucleotide having the target nucleotide sequence which is coincident with or analogous to the polynucleotide having at least one nucleotide sequence selected from SEQ ID NOS:2 to 3501; and
(iv) an output devices that shows a function obtained by the comparator.

28. A method based on a computer for determining a function of a polypeptide encoded by a polypeptide encoded by a polynucleotide having a target nucleotide sequence derived from a coryneform bacterium, comprising the following:

(i) inputting at least one nucleotide sequence information selected from SEQ ID NOS:2 to 3501, function information of a polypeptide encoded by the nucleotide sequence, and target nucleotide sequence information;
(ii) at least temporarily storing said information;
(iii) comparing the at least one nucleotide sequence information selected from SEQ ID NOS:2 to 3501 with the target nucleotide sequence information; and
(iv) determining a function of a polypeptide encoded by a polynucleotide having the target nucleotide sequence which is coincident with or analogous to the polynucleotide having at least one nucleotide sequence selected from SEQ ID NOS:2 to 3501.

29. A system based on a computer for determining a function of a polypeptide having a target amino acid sequence derived from a coryneform bacterium, comprising the following:

(i) a user input device that inputs at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001, function information based on the amino acid sequence, and target amino acid sequence information;

Table 1 (continued)

SEQ NO. (DNA)	SEQ NO. (a.a)	Initial (nt)	Terminal (nt)	ORF (bp)	db Match	Homologous gene	Identity (%)	Similarity (%)	Matched length (a.a)	Function
3418	6918	3301303	3300119	1185						
3419	6919	3301358	3301729	372	sp:THI2_CHLRE	Chlamydomonas reinhardtii thi2	42.0	76.5	119	thioredoxin ch2, M-type
3420	6920	3301755	3302996	1242	sp:CWLB_BACSU	Bacillus subtilis cwlB	51.0	75.4	196	N-acetylmuramoyl-L-alanine amidase
3421	6921	3302765	3301989	777						
3422	6922	3303435	3304475	1041						
3423	6923	3303616	3302999	618	pir:D70851	Mycobacterium tuberculosis H37Rv Rv3916c	34.4	58.5	212	hypothetical protein
3424	6924	3304787	3303636	1152	sp:YGI2_PSEPU	Pseudomonas putida ygi2	37.6	60.5	367	hypothetical protein
3425	6925	3305671	3304835	837	sp:YGI1_PSEPU	Mycobacterium tuberculosis H37Rv parB	65.0	78.0	272	partitioning or sporulation protein
3426	6926	3306532	3305864	669	sp:GIDB_ECOLI	Escherichia coli K12 gidB	36.0	64.7	153	glucose inhibited division protein B
3427	6927	3307632	3306682	951	pir:A70852	Mycobacterium tuberculosis H37Rv Rv3921c	44.7	75.4	313	hypothetical membrane protein
3428	6928	3308369	3307971	399	sp:RNPA_BACSU	Bacillus subtilis rnpA	26.8	59.4	123	ribonuclease P protein component
3429	6929	3308747	3308412	336	gp:MAU19185_1	Mycobacterium avium rpmH	83.0	93.6	47	50S ribosomal protein L34
3430	6930	3309028	3309321	294						
3431	6931	3309043	3308822	222						
3432	6932	147980	147573	408	gp:AF116184_1	Corynebacterium glutamicum panD	100.0	100.0	136	L-aspartate-alpha-decarboxylase precursor
3433	6933	268001	266154	1848	sp:LEU1_CORGL	Corynebacterium glutamicum ATCC 13032 leuA	100.0	100.0	616	2-isopropylmalate synthase
3434	6934	269068	268814	255	sp:YLEU_CORGL	Corynebacterium glutamicum (Brevibacterium flavum) ATCC 13032 orfX	100.0	100.0	85	hypothetical protein
3435	6935	270660	271691	1032	sp:DHAS_CORGL	Corynebacterium glutamicum asd	100.0	100.0	344	aspartate-semialdehyde dehydrogenase
3436	6936	446075	446521	447	gp:AF124518_1	Corynebacterium glutamicum ASO19 aroD	100.0	100.0	149	3-dehydroquinase